Intelligent System for Recommending Study Level in English Language Course Using CBR Method

Mirza Sutrisno
Faculty of Technology and Information
Budi Luhur University
Jakarta, Indonesia
mirzaindonesia@gmail.com

Utomo Budiyanto
Faculty of Technology and Information
Budi Luhur University
Jakarta, Indonesia
utomo.budiyanto@budiluhur.ac.id

***Abstract*—In the admission process, an English Course uses a level placement test. The implementation of the test encountered some problems such as slow determination of student learning levels based on the results of paper based test that are still conventional. The purpose of this research provides the recommendations for an intelligent knowledge-based system in recommending student learning levels using the Case-Based Reasoning (CBR) method. CBR is one of the method that uses the Artificial Intelligence approach and focuses on solving problems based on knowledge from the previous cases, by calculating numerical local similarity and global similarity using the nearest neighbor algorithm as the basic for the technical development of this intelligent system. The result of the study was tested for the data accuracy with the confusion matrix method by the result 100% for the accuracy. For evaluating the system systematically was using the User Acceptance Test (UAT) method with the results of the evaluation is 88% of the system meets user needs and expectations**

Keywords— Recommended Study Level, Case Based Reasoning, English Course, Nearest Neighbor

I. INTRODUCTION

English Language Courses and Training Institutions are held to meet the society needs in an effort to develop themselves in foreign language skills. In the process of enrollment, there are stages of level placement tests that are to determine the abilities of students in English. In implementing these activities, several problems were encountered, such as the slow determination of student learning levels based on the results of the paper based test, and the ineffective and efficient process of determining student learning levels.

This study builds an intelligent knowledge-based system by implementing a method in providing appropriate and accurate recommendations in determining student learning levels based on predetermined assessment criteria. Case Based Reasoning (CBR), refers to the use of experience (old cases) in decision making for new cases and the use of appropriate similarity definitions to find solutions to the case. In many situations, previous cases can be used to make further amendments to achieve new goals in decision making [1]. Case based Reasoning (CBR) collects the previous cases that are almost the same as the new case and try to modify the solution to fit the new case [2].

Case-Based Reasoning (CBR) consists of four main steps, namely:

1) *Retrieve* : retrieving the most similar cases. In this step, a searching or a calculation process is carried out from the basis of cases that have similarities.
2) *Reuse* : reusing the information and knowledge in the case base to solve the new problems. In this step, a solution of a similar case is sought from the previous conditions for the new problems.
3) *Revise* : reviewing the solution given. In this step a solution is found for a similar case in the previous conditions for the problem that occurred later.
4) *Retain* : retaining the part of previous experience to be used in solving for the future problem.

The relationship among these steps illustrated in the following Figure 1.

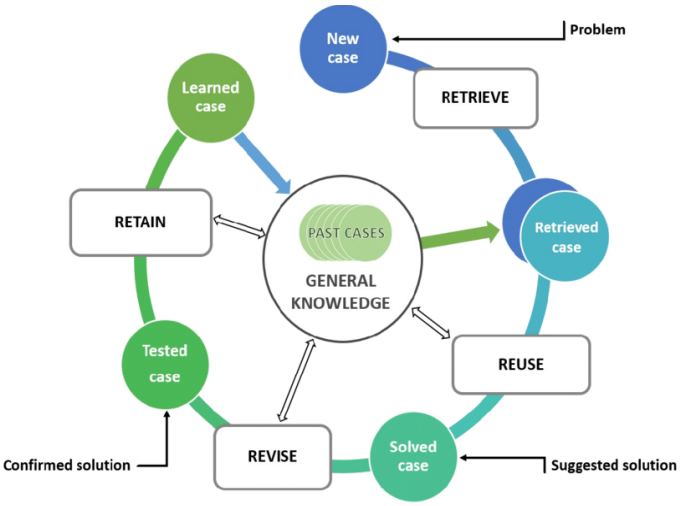

Fig. 1. Case Based Reasoning Cycle [2]

In this study the Case Based Reasoning (CBR) method is used to recommend a student learning level. The case used in this study is the score of student exam results on four assessment criteria that are as the case base (old case), the criterias are the score of reading, structure, vocabulary and listening. Every case on the case base will be calculated with the value of similarity from the case base with the new case inputted. The calculation of the similarity values through two stages of calculation, first the calculation of the local similarity value then the global similarity value. The highest global similarity value will be used as a recommendation value for a solution to determine the level of student in English language courses and training institutions.

The calculation of local similarity value is divided into two, namely numerical and symbolic local similarity. In the case of the calculation for this study, we use the numerical local similarity [3]. The computation of the local similarity

$f(S,T)$ is an equation to calculate the local numerical similarity values (1).

$$f(S,T) = 1 - \left(\frac{s-t}{R}\right) \qquad (1)$$

Where *s* is the value of the old case attribute and *t* for the value of the new case inputted and *R* as the range.

Furthermore, the calculation of global similarity is used to calculate the similarity between new cases and old cases stored on a case base (2). The method for calculating the similarity weights in global similarity is using the nearest neighbor algorithm[4]. The total of each local similarity $\sum_{i=1}^{n} F(Si,Ti)$ is multipled with the weight of each case Wi then will be devided with the total of al weight for the each case $\sum_{i=1}^{n} Wi$. It can be written:

$$(S,T) = \frac{\sum_{i=1}^{n} F(Si,Ti) * Wi}{\sum_{i=1}^{n} Wi} \qquad (2)$$

II. Some Researches on CBR

The modeling applications using CBR has been developed in various objects in the field of research, such as the selection of majors in Senior High Schools (SMA)[5]. By using the CBR method to assist students in choosing program studies in the second year of high school (SMA). The cases are used in this study included the results of intelligence tests, student interests, and scores of several subjects. Each case on a case base will be calculated for the level of similarity with the new case entered. This study resulted in a case-based computer reasoning system with the highest level of similarity value as recommendation solution to choose majors in high school (SMA).

The CBR method has also been developed in determining the feasibility of students on receiving scholarships by determining the local similarity value of numerical features and global similarity values. This study also compared the CBR method with RBR to compare the accuracy and speed of the two methods[6]. The CBR method in the world of health is a development of CBR-based systems for diagnosing dengue fever. The diagnosis process is to enter a new case compared to the old case and then calculate the similarity value using the nearest neighbor algorithm[7]. The calculation of CBR method is also implemented to calculate the similarity value between the old case and the new case using the weighted euclidean distance, looking for similarities using two distance points between the old case and the new case which is then compared with the calculation of the similarity value with the cosine similarity method[8].

III. Research Method

The study began by identifying problems in the process of determining student learning levels based on English level placement tests. Then study the relevant literature and the implementation of the methods applied. Then the system design process is carried out and continued from the data collection for the implementation and testing of the system and system evaluation and preparation of reports, as shown in Figure 2.

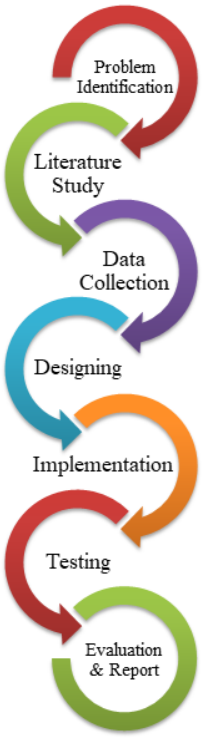

Fig. 2. Research Step

IV. Case Based Reasoning For Problem Solving

A. System Design

The Intelligent system for recommending study levels using the Case Based Reasoning (CBR) method have a system stage implemented to recommend the student learning levels. First, the students follow the placement level test by the assessment of their abilities of reading, structure, vocabulary and listening test. After the result of the test inserted then the case based reasoning system runs automatically. The general system architecture is shown Figure 3.

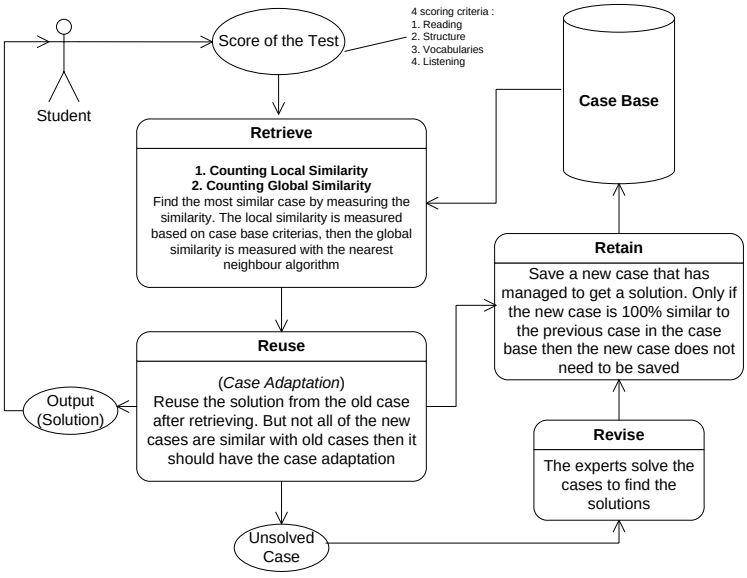

Fig. 3. System Architecture

In the initial phase, students will take the exam with the results of 4 aspects of assessment (reading, structure, vocabulary and listening) which will become the new case data. The new case data is the student identity and the assessment criteria that are used as the references for student learning levels. The next stage is processing the data that has been inputted in the CBR cycle environment including the process of retrieving, reusing, case adaptation, revision and retain. In the retrieval process, the system will search the old cases from the case base that are similar to the new case base on the criteria or the features used. The process of retrieving by using local and global similarity measurements from the target case with the value result of reading, structure, vocabulary and listening. The measurement of the local similarity is using numeric. And the measurement of the global similarity is using the nearest neighbor algorithm. The similarity value with a range between 0 to 1.

After the similar case data is found, the solution from the old case will be reused to the new case. The results of the decision as the system output will be delivered to the user as a solution through the output module. The data from the the decision on new problems and their solutions can be saved (retain) into the case base (case base) as a system learning process with the terms of the similarity of new problems less than 1.

B. Case Base Representation

The case is represented in a collection of features that the characterize of the case and the solution for handling the case. The collection of cases that have been obtained are based on the results of consultation with the expert team to determine which data can be used as a feature of the case. In the process of determining study levels, they are represented in the score of the student from the results of the reading, structure, vocabulary and listening test, which is given a weight value, as shown in Table 1.

Table I Case Base Representation

Case Base Representation			
Level	**Criteria**	**Score**	**Weight**
Foundation	Reading	24	4
	Structure	20	3
	Vocabulary	16	2
	Listening	8	1
Basic	Reading	26	4
	Structure	22	3
	Vocabulary	18	2
	Listening	10	1
Elementary	Reading	28	4
	Structure	24	3
	Vocabulary	20	2
	Listening	12	1
Intermediate	Reading	30	4
	Structure	26	3
	Vocabulary	22	2
	Listening	14	1
Advance	Reading	32	4
	Structure	28	3
	Vocabulary	24	2
	Listening	16	1

C. The Data of Student Test Result

There are 30 sample data of the students taking the test to determine the level of studying whose results will be tested for the value of system accuracy using the confusion matrix method. The following is attached in Table 2.

Table II The Data of Student Test Result

No	Reading	Structure	Vocabulary	Listening
1	28	26	20	10
2	28	24	20	12
3	28	24	24	10
4	24	26	22	12
5	26	26	20	12
6	28	24	20	12
7	26	24	20	12
8	22	24	20	12
9	26	20	20	14
10	26	22	20	14
11	24	26	20	10
12	26	24	20	10
13	24	26	20	12
14	26	26	18	10
15	26	26	18	14
16	26	24	20	14
17	26	26	22	10
18	24	24	22	14
19	30	24	20	12
20	28	24	22	12
21	22	26	18	10
22	30	22	18	10
23	26	22	18	10
24	24	22	20	8
25	24	18	10	6
26	24	20	16	8
27	28	24	20	12
28	28	24	22	14
29	20	20	16	8
30	24	18	10	6

D. Retrieval

Users (students) act as system users to take the test and the results of student tests as a reference for new cases. Then the numerical local similarity calculation process is carried out from the results. After that the global similarity uses the nearest neighbor algorithm by adding the calculation feature in the indexed case to find the same case with the highest similarity value as a solution for the recommendation of the student level. The flow of the retrieval process can be seen in Figure 4.

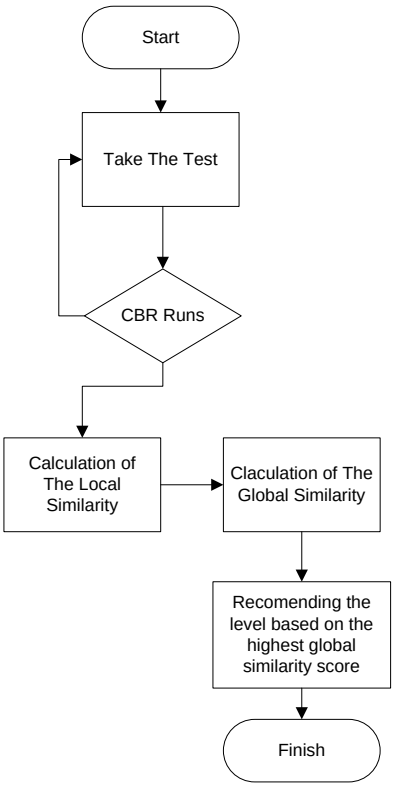

Fig. 4. Retrieval Process

E. The Calculation of Similarity

The calculation of similarity values through two stages, namely calculating local similarity based on new cases obtained from the value of student exam results on the existing case base. Then proceed with calculating the global similarity value using the nearest neighbor algorithm.

.

1. Local Similarity of foundation level:
   - Reading Score
   $$f(Si,Ti) = 1 - \left(\frac{24-24}{24}\right) = 1{,}00$$
   - Structure Score
   $$f(Si,Ti) = 1 - \left(\frac{20-18}{20}\right) = 0{,}90$$
   - Vocabulary Score
   $$f(Si,Ti) = 1 - \left(\frac{16-10}{10}\right) = 0{,}63$$
   - Listening Score
   $$f(Si,Ti) = 1 - \left(\frac{8-6}{8}\right) = 0{,}75$$

   Global similarity of foundation level
   $$Sim(S,T) = \left(\frac{(1{,}00*4)+(0{,}90*3)+(0{,}63*2)+(0{,}75*1)}{(4+3+2+1)}\right) = \mathbf{0,87}$$
2. Local similarity of foundation level:
   - Reading Score
   $$f(Si,Ti) = 1 - \left(\frac{26-24}{26}\right) = 0{,}92$$
   - Structure Score
   $$f(Si,Ti) = 1 - \left(\frac{22-18}{22}\right) = 0{,}82$$
   - Vocabulary Score
   $$f(Si,Ti) = 1 - \left(\frac{18-10}{18}\right) = 0{,}56$$
   - Listening Score
   $$f(Si,Ti) = 1 - \left(\frac{10-6}{10}\right) = 0{,}60$$

   Global similarity of basic level
   $$Sim(S,T) = \left(\frac{(0{,}92*4)+(0{,}82*3)+(0{,}56*2)+(0{,}60*1)}{(4+3+2+1)}\right) = \mathbf{0,79}$$
3. Local similarity of elementary level:
   - Reading Score
   $$f(Si,Ti) = 1 - \left(\frac{28-24}{28}\right) = 0{,}86$$
   - Structure Score
   $$f(Si,Ti) = 1 - \left(\frac{24-18}{24}\right) = 0{,}75$$
   - Vocabulary Score
   $$f(Si,Ti) = 1 - \left(\frac{20-10}{20}\right) = 0{,}50$$
   - Listening Score
   $$f(Si,Ti) = 1 - \left(\frac{12-6}{12}\right) = 0{,}50$$

   Global similarity of elementary level
   $$Sim(S,T) = \left(\frac{(0{,}86*4)+(0{,}75*3)+(0{,}50*2)+(0{,}50*1)}{(4+3+2+1)}\right) = \mathbf{0,72}$$
4. Local similarity of intermediate level:
   - Reading Score
   $$f(Si,Ti) = 1 - \left(\frac{30-24}{30}\right) = 0{,}80$$
   - Structure Score
   $$f(Si,Ti) = 1 - \left(\frac{26-18}{26}\right) = 0{,}69$$
   - Vocabulary Score
   $$f(Si,Ti) = 1 - \left(\frac{22-10}{22}\right) = 0{,}45$$
   - Listening Score
   $$f(Si,Ti) = 1 - \left(\frac{14-6}{14}\right) = 0{,}43$$

   Global similarity of Intermediate level

$$Sim(S,T) = \left(\frac{(0{,}80 * 4) + (0{,}69 * 3) + (0{,}45 * 2) + (0{,}43 * 1)}{(4+3+2+1)}\right) = \mathbf{0,66}$$

5. Local similarity of advance level:
   - Reading Score
   $$f(Si,Ti) = 1 - \left(\frac{32-24}{32}\right) = 0{,}75$$
   - Structure Score
   $$f(Si,Ti) = 1 - \left(\frac{28-18}{28}\right) = 0{,}64$$
   - Vocabulary Score
   $$f(Si,Ti) = 1 - \left(\frac{24-10}{24}\right) = 0{,}42$$
   - Listening Score
   $$f(Si,Ti) = 1 - \left(\frac{16-6}{16}\right) = 0{,}38$$

Global similarity of level advance

$$Sim(S,T) = \left(\frac{(0{,}75 * 4) + (0{,}64 * 3) + (0{,}42 * 2) + (0{,}38 * 1)}{(4+3+2+1)}\right) = \mathbf{0,61}$$

Based on the calculation above, the similarity value of foundation level is **0,87** and the basic level is **0,79** the elementary level is **0,72** and the intermediate level is **0,66** and the advanced level is **0,61**. Then the similarity level of the foundation level is greater than other similarity values, it can be concluded that the **foundation level** is the recommendation for a study level based on the highest level of similarity value.

V. RESULTS AND DISCUSSION

A. *The Testing of Accuracy*

The testing of accuracy uses the confusion matrix method, which is from 30 data of students, there are 20 students tested based on elementary levels. The following is attached in Table 3.

Tabel III Confusion Matrix Data Test

NO	READING	STRUCTURE	VOCABULARY	LISTENING	RESULT
1	28	26	20	10	0,958
2	28	24	20	12	1
3	28	24	24	10	0,943
4	24	26	22	12	0,938
5	26	26	20	12	0,946
6	28	24	20	12	1
7	26	24	20	12	0,971
8	22	24	20	12	0,914
9	26	20	20	14	0,938
10	26	22	20	14	0,930
11	24	26	20	10	0,901
12	26	24	20	10	0,955
13	24	26	20	12	0,918
14	26	26	18	10	0,960
15	26	26	18	14	0,993
16	26	24	20	14	0,988
17	26	26	22	10	0,943
18	24	24	22	14	0,906
19	30	24	20	12	0,971
20	28	24	22	12	0,980

From the data above, it can be classified into the table of confusion matrix calculations in Table 4.

Table IV. Confusion Matrix Classification

Actual Value		Prediction	
		Positive of study level (true).	Negative of study level (difference)
Actual Value	Positive of study level (true).	20 (TP)	0 (FP)
	Negative of study level (difference)	0 (FP)	0 (TN)

The Confusion matrix classification table shows 20 cases that get a global similarity value of more than 0.9 for the elementary study level as the expertist actual value. The value results on the actual data has the same predictions. So that the level of accuracy can be calculated:

$$\textbf{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn}\ X\ 100\% = \frac{20+0}{20+0+0+0}\ X\ 100\% = \mathbf{100}\%$$

The calculation of the results show that the calculation of the global similarity weights with the nearest neighbor algorithm is able to recognize cases of appropriateness of student learning levels correctly with **100%** accuracy**.**

B. *User Acceptance Test* (UAT)

For the UAT Testing, samples are taken by convenience sampling method where samples are selected according to the wishes of the researcher with the reason of the availability of members or easily obtained by taking first batch of the exam participants by taking a sample of 30 students to fill out the online questionnaire. The results of these tests with the likert scale are very good with a user acceptance rate of 88%. Enclosed in Figure 5.

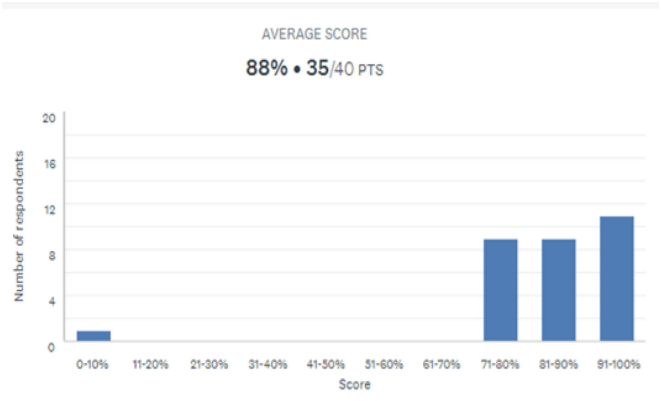

Fig. 5. UAT Questionaire Result

VI. CONCLUSION

Based on the description, the explanation and the testing that has been done, there are some conclusions as follows :

1. This study produces an intelligent system model for recommending study level. The system

model is independently developed according to the requirements.

2. The recommended study level based on the student test scores includes four aspects of the assessment criteria, such as reading, structure, vocabulary, and listening which were tested in multiple choices with 50 questions. Based on the test scores, the CBR (Case Based Reasoning) method is used to calculate the numerical local similarity and global similarity, as the basic for determining study level. There are five recommended levels namely Foundation, Basic, Elementary, Intermediate, Advance .
3. Based on the results of testing the proposed model, it is known that the level of student learning is dominant at the elementary level and based on accuracy testing that the recommended learning level is considered to be very effective by calculating the accuracy of the data using the confusion matrix method and the final results of the test with the accuracy value of 100%.
4. The software testing involved 30 students who took the learning level determination test by distributing online questionnaires with the results of the test helping to speed up the process of determining the level of learning and overall system acceptance very well with an acceptance percentage of 88%.

REFERENCE

[1] D. Wang, Z. Li, N. Dey, A. S. Ashour, R. S. Sherratt, and F. Shi, "Case-Based Reasoning for Product Style Construction and Fuzzy Analytic Hierarchy Process Evaluation Modeling Using Consumers Linguistic Variables," *IEEE Access*, vol. 5, pp. 4900–4912, 2017.

[2] A. Agnar and E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *AI Communication*, vol. 7, pp. 39–59, 1994.

[3] S. Guessoum, M. T. Laskri, and J. Lieber, "RespiDiag: A case-based reasoning system for the diagnosis of chronic obstructive pulmonary disease," *Expert Syst. Appl.*, vol. 41, no. 2, pp. 267–273, 2014.

[4] L. S. Ong *et al.*, "The Colorectal Cancer Recurrence Support (CARES) System," *Artif. Intell. Med.*, vol. 11, no. 3, pp. 175–188, 1997.

[5] S. Mulyana, S. Hartati, R. Wardoyo, and E. Winarko, "Case-Based Reasoning for Selecting Study Program in Senior High School," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 4, pp. 136–140, 2015.

[6] F. Tempola and S. Do Abdullah, "Case-Based Reasoning (CBR) untuk Penentuan Kelayakan Mahasiswa Penerima Beasiswa," *Jurnal Protek*, vol. 05, no. 2, pp. 1–5, 2018.

[7] R. Adawiyah, "Case Based Reasoning Untuk Diagnosis Penyakit Demam BerdarahAdawiyah, R. (2018) 'Case Based Reasoning Untuk Diagnosis Penyakit Demam Berdarah,' Intensif, 1(1), hal. 63. doi: 10.29407/intensif.v1i1.544.," *Intensif*, vol. 1, no. 1, p. 63, 2018.

[8] H. M. Jumasa, S. Fauziati, and A. E. Permanasari, "Penerapan Case-Based Reasoning Dalam Menentukan Similarity Berdasarkan Kesesuaian Lahan Kelapa Sawit," pp. 306–313, 2017.